Vol. 6, Doc. 44a. "The Principal Ideas of the Theory of Relativity" [p. 1]

[after December 1916][1]

Ask an intelligent man who is not a scholar what space and time are, and he will perhaps answer as follows. If we imagine all physical things, all stars, all light taken out of the universe, what then remains is something like a giant vessel without walls called "space." With respect to what is happening in the world, it plays the same role as the stage in a theater performance. In this space, in this vessel without walls, there is an eternally uniformly occurring tick-tock that, however, only ghosts, but those everywhere can hear; that is "time." Most natural scientists, up to the present, had this conception about the essence of space and time, even though they did not phrase it in such naive terms as we just did for the sake of simplicity.

Based upon this conception one is inclined to make immediate sense of the following statements. Two eruptions of Mount Vesuvius occur at different times but at the same place (that is, at the crater of Mount Vesuvius). The lighting-up of two distant "new stars" occurs at the same time but at different locations. It has long been known that statements of the first kind (on equi-locality) make no sense. Indeed, the earth rotates about its axis, moves around the sun, and furthermore, moves together with the sun toward the constellation of Hercules. Therefore, one cannot seriously claim that the two eruptions of Mount Vesuvius occurred at the same location in the universe. From this example it is seen easily that no sense can be attributed to such statements about equi-locality. We can only say: the two eruptions of Mount Vesuvius occur at the same place *relative to the earth*. The earth [p. 2] plays the role of a "body of reference"; topographic statements make sense only if they are referred to a body of reference.

In contrast, statements about simultaneity, however, or time in general, seem to make sense independent of any body of reference. Without much reflection, one would be inclined to declare any human insane who claims that the statement of two stars lighting up simultaneously makes no sense unless a body of reference is pointed out relative to which simultaneity occurs. And yet, the convincing force of experimental facts has forced science to make that claim. How did this happen?

Experiences about the propagation of light led to this strange result. Based upon many experiments, physicists became convinced that light propagates through empty space at a speed of $c = 300{,}000$ kilometers per second, entirely independent of the velocity of the body that emits this light. Imagine a ray of light sent by the

sun in a distinct direction. According to the law just stated, this ray travels a distance of c per second. Now imagine the sun later hurls a body into space such that it flies with a velocity of 1,000 kilometers per second in the same direction as the ray of light. This is easy to imagine. We now can similarly imagine this project-
[p. 3] ed body as an alternative body of reference and ask ourselves: what is the propagation velocity of light in the judgment of an observer who does not sit on the sun but rather on the projected body? The answer seems simple. When the hurled body runs after the light at 1,000 kilometers per second, the ray of light advances against it by only 299,000 kilometers per second. The same situation would prevail if the ray of light were not sent by the sun but rather by the projected body, because we know that the velocity of light does not depend upon the state of motion of the light source.

This result raises suspicion. Should light when judged from the projected body really propagate differently than when judged from the sun? Should the laws of the propagation of light depend upon the state of motion of the body of reference? Then there would have to exist in the universe something like absolute rest, because one could argue like this. Relative to arbitrarily moving bodies of reference (here the projected body), light propagates with a velocity different from c that depends on the direction. Then there would exist a body of reference of a very distinct state of motion for which light propagates in all directions with the same velocity c. Such bodies of reference could justifiably be said to be at absolute rest (in our example, the sun). Does such absolute rest really exist in a physical sense? Do the laws of nature really depend upon the state of motion of the observer, i.e., on the system of reference, as suggested by the argument above about the propagation of light?

Experience contradicts this. When we travel in a railroad car that is free of vibra-
[p. 4] tion we do not notice the motion of the car. All experiments of physics in this car show the same success as found in a house resting on earth. The physical experiments done on earth show no effects of the movement of the whole earth with all its objects on it. In general: *the laws of nature are independent of the state of motion of the body of reference.* This statement is called, for short, the "principle of relativity." But we believed that the consideration given above forced us to conclude that with respect to the law of light propagation the principle of relativity was *not* valid; what is the truth? More than 30 years ago Michelson, an American, proved through his famous optical experiment that the principle of relativity would be also valid in a case where theory allowed to predict the influence of the earth's movement upon the experiment.[2]

Therefore, the considerations given above must have contained an error. The law of light propagation is the same, whether the sun or the projected body is chosen as the body of reference. The same ray of light travels at 300,000 kilometers per sec-

ond relative to the sun and also relative to the body projected at 1,000 kilometers per second. If this appears impossible, the reason is that the hypothesis of the absolute character of time is false. One second of time as judged from the sun is not equal to one second of time as seen from the projected body.

There is no audible tick-tock everywhere in the world that could be considered as time. If physics wants to use time, it first has to define it. In this endeavor it is apparent that this definition necessarily requires a body of reference, and that this definition makes sense only with respect to this chosen body of reference. It turns [p. 5] out that one can define time relative to this body of reference such that the law of the propagation of light is obeyed relative to it. This definition of time can be realized for bodies of reference in any state of motion. But it turns out that the times of differently moving bodies of reference do not coincide. A more detailed justification of this is found in my popular book about the theory of relativity.[3] If two events occurring at different locations are judged simultaneous from a body of reference, then they are not judged so from a body of reference that is moving relative to it.

Before I continue in this train of thought, I have to say something about the role the body of reference plays in Galileo's and Newton's mechanics. In particular I have to point out that the development of science knows only of a build-up, not of a tearing-down. If a generation cannot build upon the achievements of its predecessors, then there is no science proper. It would be sad if the theory of relativity would have to topple the previous mechanics, somewhat like one tyrant toppling the other. The theory of relativity is nothing but a step further in the centuries-old development of our natural sciences, which preserves and deepens previously found connections and adds new ones. The theory of relativity does not topple Newton's and Maxwell's theories, just as the League of Nations[4] does not annihilate the states that join it. They will have to accept some modifications of their laws but thereby gain higher security.—

In everyday life we mostly use the surface of the earth as a body of reference whose individual points can be repeatedly identified. Mathematical physics chooses as a body of reference (coordinate system) three mutually orthogonal straight rods originating from one point. The position of a point relative to this system of rods is described by three numbers (coordinates) that can be obtained by measuring with rigid rods (measuring rods). For this procedure it is assumed that the laws of orienting rigid bodies are correctly described by Euclidean geometry. All state- [p. 6] ments of location made by physics hitherto are based upon this assumption. Wherever a point may be located, one can always think of the system of rods and the procedures of measurement to be completed such that eventually they reach up to the point under consideration. This must be imagined like scaffolding at a

construction site where one can reach every little spire and corner, no matter how tall the building. In physics it is not even necessary that this scaffolding exists in reality, provided one can imagine it being built indirectly (with rays of light etc.).

The fundamental mechanical laws of Galileo and Newton are such that they do not claim validity relative to arbitrarily moving bodies of reference but only relative to those of suitably chosen states of motion. Bodies of reference that are admissible in mechanics are called "inertial systems." Now, there is a theorem of mechanics: if the body of reference K is an inertial system, then any other body of reference moving uniformly in a straight line and without rotation relative to K is also an inertial system. Phrased more simply: if the laws of mechanics hold relative to the surface of the earth, then they also hold relative to a uniformly moving railroad car that is taken as a body of reference.

What has been said previously about light can now be summarized in a simple formula: relative to every inertial system—given the correct definition of time—the theorem of the constancy of the speed of light in empty space holds true. More generally, one can express as a theorem of manifold experience: the laws of nature are the same in all inertial systems. This theorem is called "principle of special relativity."

That this theorem implies a novel method of research can be understood in the following manner. Assume the universe or the individual events that constitute it have been described relative to *one* inertial system, then the course of events *seen from a different inertial system* is a different one, but nevertheless is also completely determined. The Dutch mathematical physicist calculated the general rules that
[p. 7] allow one to transform location and time from one inertial system into another.[5] Obviously, in this manner one can not only transform individual events but also mathematically formulated laws of nature. The principle of special relativity demands of these laws that they do not change under such transformation. If they do not have this property, then they have to be rejected by the principle of special relativity. The laws of nature must be adapted to the principle of special relativity.

The need to modify Newtonian mechanics first emerged during the investigations that dealt with extremely fast movements, or more precisely, when motions approached a speed that could no longer be treated as negligibly small compared to the speed of light. Furthermore, it turned out that the inertia of a body is not a characteristic constant of that body, but that inertia rather depends upon the energy content. *Mass and energy in essence are identical.*

31. "Fundamental Ideas and Methods of the Theory of Relativity, Presented in Their Development"

[after 22 January 1920][1]

Fundamental Ideas and Methods of the Theory of Relativity, Presented in Its Development.[2] [p. 1]

On the occasion of ⟨confirmation⟩ the finding of the gravitational curvature of light rays by the British expedition that was sent to observe the eclipse of the sun, I have been urged by many to give a brief description to non-mathematicians of the theory and its development.[3] I do this the more gladly as there is a certain danger that the—unfortunately—rather complicated mathematical form of the theory threatens to overshadow its simple ⟨and natural⟩ physical content. This mathematical form is merely a means, while the essence of the theory is definitely the consequent execution of some ⟨less taken from physical experience⟩ ⟨general⟩ simple principles to which physical experience has led us. The theory is not at all, as many believe, the result of daring mathematical speculation.[4]

I. The Special Theory of Relativity

The special theory of relativity is nothing but a contradiction-free amalgamation of the results of Maxwell-Lorentz electrodynamics and those of classical mechanics. This will become clearly obvious from the following schematic sketch of its development.

1. *The Light Ether*

The investigation of interference and diffraction phenomena of light during the first half of the 19th century showed that ⟨light as⟩ Huygens's opinion of the undulatory nature of light was correct versus Newton's theory of emission. Since physicists then were convinced that all processes in nature have to be interpreted mechanically, they viewed light as the undulatory motion of a hypothetical substance called "ether." The introduction of a specific medium of light aside from other matter was

necessary because light propagates with a well-defined velocity c independent of the wavelength through (empty) space that is free of matter in the ordinary sense.

2. *Light Ether and the Movement of Matter*

Soon, one pondered the question of what kind of substance this ether should be perceived of. Is it akin to a fluid or to a solid? The fact of polarizability of light led to the conception that the oscillations of light are transversal oscillations, that is, oscillations as they occur only in solid but not in fluid bodies. This led to the conception of the ether as a kind of solid body, i.e., a substance that resists changes of shape or relative motions of its parts by virtue of a lively elastic resistance. It seemed to behave like a quasi-rigid body that penetrates all matter.

The fundamentally important Fizeau experiment (1851),[5] which tried to answer the question if moving matter carries the light ether contained in the same volume with itself, also led to the same conception. The consideration was the following one. A fluid at rest propagates light with velocity V ($V = \frac{c}{n}$, where n is the index of refraction). If the fluid flows with velocity v from left to right through a pipe, *and the fluid carries its light ether with it*, then the light ray sent through the
[p. 2] fluid from left to right will also have propagation velocity V relative to the *flowing* fluid. Its propagation velocity relative to the pipe will then, due to the addition theorem of velocities, be larger by v than it is relative to the fluid, therefore

$$V' = V + \mathrm{v}. \tag{1}$$

But the Fizeau experiment did *not* confirm this result of the consideration. Empirically, one found the correctness of the Fresnel formula[6]

$$V' = V + \left(1 - \frac{1}{n^2}\right)\mathrm{v}_\circ \tag{2}$$

Thus it is shown that the propagation velocity of light is less strongly influenced by the movement of matter than the previous consideration made us expect. For fluids that do not refract light ($n = 1$), formula (2) even yields

$$V' = V, \tag{2a}$$

i.e., the movement of a non-light-refracting fluid has no influence upon the propagation velocity of the light penetrating it.

This result forced the conception that the light ether does not partake at all in the movement of matter and that the influence of moving matter upon light, as given in (2), is not to be explained by a movement of the light ether but rather in another more indirect manner. This other explanation was later given by H. A. Lorentz in a very complete and satisfying manner by maintaining the hypothesis of a fixed—or as one can also say—light ether at rest.[7]

In passing, we mention that the fact of the aberration of light (discovered by Bradley in 1727) could only be satisfactorily explained if one assumed that the ether at the surface of the earth does not partake in the movement of the earth around the sun.[8]

3. *The Lorentz Theory*

The theory of the light ether at rest became completely successful by the pioneering research of H. A. Lorentz (1895),[9] who simplified and deepened the Maxwellian theory and simultaneously brought it into agreement with all electromagnetic and optical results known at that time. His theory rests on the following basis:

a. In matter, too, only the ether (but not matter) is the base of the electromagnetic field.
b. Matter is electromagnetically effective only as the carrier of electric masses that are movable, together with it (when matter moves) and relative to it (when there is electric current, a changing electric polarization, a magnetization).
c. The Maxwellian equations for vacuum are valid everywhere (also inside matter) in a coordinate system relative to which the light ether is at rest.

Between Maxwell's and Lorentz's time, a slow and important change in fundamental concepts of theoretical physics took place that we cannot ignore. Maxwell himself still clung to the conception that all physical events have to be interpreted in terms of mechanics. But his efforts, and those of other important theoreticians, to devise a mechanical model of electromagnetic phenomena in the ether did not meet with success. Poincaré pointed out that even if the construction of such a picture were accomplished, it would not be a decisive success because such a picture would only be one in an infinite number of possible ones which, in principle, would [p. 3]

be equally justified.[10] Gradually, one got used to viewing electromagnetic fields as ultimate physical realities that do not require a reduction to mechanical elements. Today the mechanical interpretation of the electromagnetic field appears superfluous to the expert, because, under the impression of many failures, we became used to resignation in this respect, and furthermore, because in the latest state of theoretical development we have abandoned the opinion that mechanical processes are simpler or better known and can conceptually be grasped more easily than physical process of another kind.—

The ⟨Lorentz⟩ ether consequently has lost during this period something of its corporeality. While it was an elastic solid body in the past, it now was only a carrier of electromagnetic fields. One can say of the ether: its only *mechanical* property was its immovability.

I do not want to go into the details of the important confirmations of Lorentz's theory except to note that even today there is no other theory of electromagnetic processes comparable to it as far as truth content and internal consistency are concerned.

4. ⟨*The Principle of Special Relativity*⟩ *The Ether Wind*

One fatal aspect, however, seemed to cling to Lorentz's theory—namely, the light ether at rest which, more or less, is a materialization of absolute space as Newton introduced it. This is due to the following pressing consideration. The earth moves on its course around the sun with a velocity of 30 km/sec. Whatever the state of motion of the sun against the ether might have been, there certainly had to be a considerable velocity of the earth relative to the medium of light during part of the year. Alternatively, this can be phrased as follows: the ether flows by with very considerable speed (of at least 30 km) while we sit on this earth, and furthermore it flows
[p. 4] through us and through earth (ether wind). While the blowing of this wind cannot be felt directly, it should be noticeable at least indirectly in electromagnetic and optical experiments. One would expect, for example, that the speed of light in a vacuum in the direction of the ether wind amounts to $c + \mathrm{v}$, in the opposite direction $c - \mathrm{v}$ when v is the velocity of the earth relative to the ether. In general, one would expect with optical and electromagnetic experiments that the orientation of the apparatus relative to the direction of the ether wind should show some influence.

Naturally, the experimental confirmation of the ⟨directed⟩ existence of the ether wind tempted physicists immensely. Most diverse apparatuses were investigated whose total orientation relative to the direction of the ether wind could be changed

and which, due to the daily rotation of the earth, would change by itself. *But no effect ever could be verified that would allow for an interpretation as the action of the ether wind.*

H. A. Lorentz himself ⟨felt and⟩ saw clearly that this negative outcome was a fact of the highest significance, and his theory had to deal with it. With penetrating calculations he proved that, according to his theory, the ether wind could not have any experimentally verifiable effect in nearly all actually performable and performed tests. This proof was based upon the fact that the velocity v of the ether wind is so small compared to the speed of light c that the quantity $\left(\frac{\mathrm{v}}{c}\right)^2$ becomes a practically vanishing ratio (order of magnitude one-hundredth of a million, 10^{-8}). There was only *one* experiment, namely the famous one by Michelson (1881, repeated in 1887 with even higher accuracy by Michelson and Morley)[11] and its precision was so high that, according to Lorentz's theory, it would definitely have been sufficient to demonstrate the ether wind.

It is well worthwhile to consider the gist of this experiment even though its proper description is not necessary for our purposes. Let AB be a rod of length l. Let a ray of light go from A to B, reflect it in B so that it comes back to A. Light requires a time of $\frac{2l}{c}$ to traverse this distance if the rod is at rest relative to the ether. But if the rod is moved relative to the ether with a velocity v, one obtains a time different from $\frac{2l}{c}$; different (Δ_1, Δ_2, respectively) depending on the rod being oriented in the direction of the ether wind or perpendicular to it. From elementary considerations, one obtains

$$\Delta_1 = \frac{2l}{c} \cdot \frac{1}{1 - \frac{\mathrm{v}^2}{c^2}}$$

$$\Delta_2 = \frac{2l}{c} \cdot \frac{1}{\sqrt{1 - \frac{\mathrm{v}^2}{c^2}}} .$$

The time difference $\Delta_1 - \Delta_2$ in Michelson's interference experiment should have caused a shift in the interference pattern, viewed via telescope, when the entire apparatus was rotated by 90° about a vertical axis. The result did not occur; even this ether-wind-weather vane of utmost sensitivity did not feel the ether wind. [p. 5]

Lorentz and Fitzgerald also found an explanation of this puzzle[12] and—as we shall see—the essentially correct explanation. They said: After all, it might be possible that the ether wind also acts upon the rod AB, and in general upon all solid bodies. Let us assume all rigid bodies when moved with velocity v relative to the ether suffer a shortening in the direction of the movement (i.e., in the direction of the ether wind) in the ratio $1 : \sqrt{1 - \frac{v^2}{c^2}}$, without change in their transversal dimensions. Then, one would have to substitute in the formula for Δ_1 the rod length $l\sqrt{1 - \frac{v^2}{c^2}}$ for the rod length l, and the times Δ_1 and Δ_2 would become equal.

This hypothesis formally satisfies the factual situation, but the mind remains ⟨quite⟩ unsatisfied with all this.[13] Should ⟨God⟩ nature really have put us into an ether storm, should, on the other hand, ⟨he⟩ she have arranged nature's law precisely so that we can never notice the storm?[14] As a matter of fact, the truth is we ourselves were the ones who did invent the ether together with the ether wind. Could we not just free ourselves from the entire difficulty by relegating ether wind and ether into the realm of imagination? We shall immediately see that the difficulty is more deeply rooted.

5. *A More Precise Formulation of the Problem*

a. According to Lorentz's theory, there exists a coordinate system K (body of reference) of a specific state of motion, relative to which every ray of light through spaces free of matter (vacuum) propagates at velocity c. (This follows necessarily from the Maxwell-Lorentz equations of the electromagnetic field.) In the following we shall remember this statement and call it the "principle of the constancy of the speed of light."[15]

b. If one introduces next to the coordinate system K a second one K' that moves in a straight line, uniformly and free of rotation relative to K (it corresponds to earth in the previous consideration), then intuitive perception seems to suggest for this system: light rays in vacuum with velocity c in K do *not* have velocity c in K' but are rather direction-dependent in their velocity.

c. On the other hand, one has to accept as an expression of experience (e.g. from the Michelson experiment): the systems K' and K are equivalent with respect

to the law of light propagation. Experience shows at least that also with respect to K', all directions are optically equivalent.

There is a contradiction between conclusion (b) and experimental finding (c). The special theory of relativity resolves this contradiction in the following manner. [p. 6]
It retains the statements (a) and (c) and demonstrates the falsity of conclusion (b).

Before we resolve the contradiction, we generalize statement (c)—based on the fact that terrestrial experiments alone can in no way demonstrate the progressive movement of the earth—by stating the theorem that is also valid in Newtonian mechanics: If one starts out with an "admissible" coordinate system K (an inertial system of classical mechanics), then every system K' that is moving uniformly and is non-rotating relative to K is equivalent to K, and the laws of nature are exactly the same in both K and K' (principle of special relativity).

6. *Epistemological Note*

A theory has physical content only if the quantities tied together in equations make physical sense, i.e., it has to be stipulated precisely how these quantities ⟨in nature⟩ can be determined ⟨or calculated⟩ from results of direct measurements. A theory says nothing about nature if the stipulations of relations between mathematical quantities of the theory and the results of measurements are missing.[16]

If I talk, for example, about the length l of a rod, then I want to say that the basic unit of measurement can be laid out l times next to it. When I say the Cartesian coordinates of a point are x, y, z, then I claim, first, that from rods of equal lengths, when their ends are suitably joined, a cubic lattice can be constructed, such that (rigid) rods behave like distances of Euclidean geometry. Further, I take such a lattice ⟨system⟩ in a certain state of motion as given, define a point as the origin and the Cartesian coordinates in a certain way as countings obtainable through this lattice.

When I speak of the temporal duration of a process, I mean, as a physicist, the number of periods that have elapsed on the clock that is used as a basis, counting from the beginning to the end of the process. (Preliminary, imprecise definition.)

It is true that a physical theory can have mathematical quantities that do not satisfy these conditions; they then are auxiliary parameters which ⟨per se⟩ have no other meaning than to simplify the expression of laws in certain cases or to unify the theoretical framework. We shall talk later about important cases of this kind.

7. *The Relativity of Simultaneity*[17]

The idea which led out of the dilemma explained under (5) is connected to the following question. Let A and B be two points of the coordinate system K. Let a sudden event take place in each one of these points. We also state that these two events occur *simultaneously*. Does this statement have a definite meaning? If yes, what is this meaning?

[p. 7] At first, everybody believes that the statement of simultaneity would make immediate sense; but this certainty rests on a deep-rooted prejudice suggested by experience. ⟨When we want to clarify what we understand under the term simultaneity of both events.⟩ Because for most purposes it is sufficient to agree that: two events are simultaneous if one observer sees them simultaneously. For if the locations of both events are not too far distant from each other, then the simultaneous viewing is nearly independent of the location of the observer; light has a speed of propagation that in most experiences is equivalent to a velocity of infinite magnitude. Light practically provides instantaneous signals; it embodies simultaneity with sufficient precision for everyday experience. For this reason alone we are inclined to connect immediate sense with simultaneity of spatially distant events.

On the other hand, it would be impossible to chronologically link the processes at A with those at B if there were no physical interaction between the locations A and B. It is therefore clear that we can get a physical definition of simultaneity only if our definition uses processes that establish a connection between the two locations considered. For such processes we select light signals in a vacuum because we are especially well informed about the law of light propagation, as was explained above. We define simultaneity on the basis of the principle of the constancy of the speed of light.

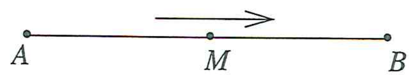
Let A and B be two points of the inertial system K, say at the end points of a rod, at rest relative to K, and M be the midpoint of the rod. A light signal shall be sent from M in all directions. The principle of the constancy of the speed of light forces us to state that the arrival ⟨time⟩ of the light signal in A is simultaneous with the arrival in B. With this we have won a physically meaningful definition of simultaneity.[1][18]

However, it is to be well noted that this definition uses the system of reference K. We do not know if two distant events that are simultaneous in reference to K

[1] The following is necessary in order to keep the definition free of contradiction. If the ⟨pair of events⟩ (α, β) and (α, γ) are pairs of *simultaneous* events with respect to K, then β and γ too are simultaneous. If this were not the case, the principle of the constancy of the speed of light could not be maintained.

are also simultaneous in reference to a second inertial system K' when K' is in motion relative to K. Indeed, a simple analysis shows that according to our principle this is not at all the case.

Fig. 1



Let the rod $A-B$ be moving with constant velocity in the direction $A-B$ relative to the inertial system. The rod is then at rest relative to its own inertial system K' (the co-moving system). If now, again, a light signal is sent from M, then—from what has been said above—the events of arrival in A and B are simultaneous when seen from K' (i.e., from the rod). Due to the movement of the rod when seen from the (not co-moving) system K, the light ray $M-B$ has to traverse a longer path. Therefore, it takes longer than the light ray $M-A$. The arrival in B occurs—when judged from K—later than the arrival in A, while—when judged from K'—they arrive simultaneously. [p. 8]

In this consideration, the validity of the principle of the constancy of the speed of light has been assumed—in agreement with the principle of special relativity—for both systems K *and* K'.

When there are identical clocks on the rod, resting in A and B, respectively, i.e., at rest relative to K', then let them be adjusted such that their hands show simultaneity relative to K'. Clocks arranged in this manner and at rest relative to K' shall be called "synchronous" (i.e., identically adjusted) relative to K'. The totality of identically constructed and identically adjusted clocks, at rest relative to K', shows the "time of system K'." According to this, every admissible coordinate system (inertial system) has its own time.

8. *The Relativity of Length*

For the length of a rod AB, moving uniformly relative to K and at rest relative to K', one can give the following two definitions:

a. Direct measurement of length by repeated laying off of a unit measuring rod, at rest relative to K', along AB.
b. Determination of those points A^* and B^* of system K, where the ends A and B of the rod are located at a certain time t of system K; where the measurement of the distance $A^* B^*$ is achieved by the repeated laying off of a unit

measuring rod that is at rest relative to K.[2][19]

It is obvious that the results l' and l of the two entirely different processes of measurement do not need to be the same. (Relativity of lengths.)

It is now easily seen that the relativity of times and lengths does not justify the consequence drawn at section 5b. The dilemma expounded in sec. 5 is herewith resolved.

9. *Galilei Transformation*

Description in physics always uses a coordinate system (body of reference) upon which all processes are referred. Whatever occurs is composed of "pointlike
[p. 9] events," each one determined chronologically and spatially relative to the coordinate system through three spatial coordinates x, y, z and a time value t.

On the other hand, we know that in classical mechanics there is already an infinite manifold of admissible coordinate systems (inertial systems), all completely equivalent for the description of nature. If x, y, z, t are the space-time coordinates of a pointlike event in reference to system K, and x', y', z', t' are the coordinates of the same event in reference to a system K' that moves with velocity v relative to K, then it is clear that, with a given orientation and location of K' relative to K, the primed coordinates must be completely determined by the unprimed ones (coordinate transformation).

Classical mechanics implicitly assumed the absolute character of times and lengths. Consequently, for the two coordinate systems whose relative position and orientation is sketched in the attached figure, one had to assume the transformation:

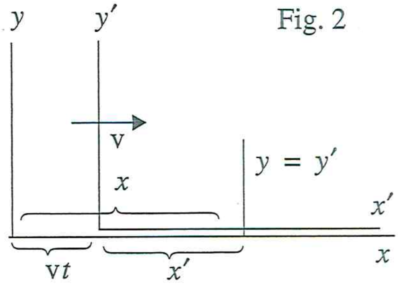


Fig. 2

[2] The unit measuring rods in (a) and (b) shall be equal to each other when compared at relative rest to each other. It shall be noted here that this condition of equality (also for clocks) is a lasting one, independent of the past history of motion and an essential precondition of the entire theory.

$$\left.\begin{aligned} x' &= x - \mathrm{v}t \\ y' &= y \\ z' &= z \\ t' &= t \end{aligned}\right\} . \qquad (3)$$

This transformation, which is called the "Galilei-transformation," is in an important relation with the Newtonian equations of motion in mechanics. If one uses in those equations the variables x', y', z', t' according to (3) instead of the variables x, y, z, t, one obtains in those new coordinates equations of exactly the same form. One says: the equations of classical mechanics are covariant under Galilei transformations. This is the analytical expression of the principle of special relativity in classical mechanics.

However, the Galilei transformation with its absolute time $(t = t')$ cannot, according to previous considerations, do justice to the actual behavior of measuring rods and clocks. For a light ray propagating along the positive x-axis according to

$$x = ct,$$

one would get relative to K', from (3)

$$x' = (c - \mathrm{v})t',$$

in contradiction to the requirement that the principle of the constancy of the speed of light must also hold relative to K'.

10. *The Lorentz Transformation*[20] [p. 10]

The intuitive reasoning which led to the Galilei transformation (3) is, according to our previous analysis, not sound. Because, contrary to the fourth equation in (3), we already recognized the relativity of simultaneity. Furthermore, if we rigorously want to interpret fig. 2, we have to add that the sketch has to hold for a specific time value t of the unprimed ("resting") system K. We cannot know if the coordinate of length x', as seen from K', also equals x' when seen from K (relativity of length). Consequently, the foundation of the first equation in (3) falls, and, in analogy, for the second and third equations as well.

Now, in order to get usable transformation equations in place of (3), one only has to satisfy the condition that one and the same light ray has to have speed c relative

to K as well as relative to K'. A spherical wave propagating from the origin of the coordinate system satisfies (under suitable choice) the equation

$$r = ct,$$

where we put according to the Pythagorean theorem

$$r^2 = x^2 + y^2 + z^2 .$$

Squaring the equation above we can also write

$$x^2 + y^2 + z^2 - c^2 t^2 = 0 . \tag{4}$$

Furthermore, since according to the principle of relativity the propagation of light must be the same relative to K' as it is relative to K, the *same* process of propagation relative to K' must also be described by a spherical wave of propagation velocity c. Therefore, the transformation we are looking for must be such that equation (4) and also the equation

$$x'^2 + y'^2 + z'^2 - c^2 t'^2 = 0 \tag{4a}$$

⟨are valid and⟩ mutually require each other. In essence, this condition determines the transformation of the space-time coordinates. For the case of coordinate systems oriented as sketched in fig. 2, one arrives—in a manner not to be explicated here—at the so-called Lorentz transformation

$$\left.\begin{aligned} x' &= \frac{x - \mathrm{v}t}{\sqrt{1 - \frac{\mathrm{v}^2}{c^2}}} \\ y' &= y \\ z' &= z \\ t' &= \frac{t - \frac{\mathrm{v}}{c^2}x}{\sqrt{1 - \frac{\mathrm{v}^2}{c^2}}} \end{aligned}\right\} . \tag{5}$$

Indeed, if one substitutes the coordinates x, y, z, t, using (5), for the coordinates x', y', z', t' in (4a), then, after a simple, straightforward calculation, one gets equations (4). [p. 11]

In a somewhat simpler manner, the Lorentz transformation can be characterized as implying the (identical) validity of the equation

$$x^2 + y^2 + z^2 - c^2t^2 = x'^2 + y'^2 + z'^2 - c^2t'^2. \tag{6}$$

As we shall see, this is the basis for the important formal progress the special theory of relativity has made through Minkowski.

For considerations following later, one should note that the Lorentz transformation equations (6) not only hold for the coordinates of pointlike events but also for the differences of corresponding coordinates of two pointlike events, as can easily be derived from equations (5). If the coordinate differences are infinitesimally small, meaning if the events are spatially and chronologically infinitely close, one obtains for these differences $(dx, dy, dz, dt, and, dx', dy', dz', dt')$ the equation

$$dx^2 + dy^2 + dz^2 - c^2dt^2 = dx'^2 + dy'^2 + dz'^2 - c^2dt'^2. \tag{6a}$$ {1}

11. *Physical Content of the Lorentz Transformation*

The equations (5) imply very specific physical statements about the behavior of measuring rods and clocks because the coordinates are physically defined quantities that can be measured with measuring rods and clocks. Therefore, it is an untenable view to believe that the replacement of the Galilei transformation by the Lorentz transformation constitutes a merely conventional or formal act.[21]

For example, consider the points $x' = 0$ and $x' = l$ at a certain time, e.g., $t = 0$ from system K. The first of the equations (5) yields for this

$$x = 0 \text{ and } x = l\sqrt{1 - \frac{v^2}{c^2}}, \text{ respectively.}$$

A rod of length l at rest, when moved lengthwise with velocity v, then, due to its movement relative to the coordinate system that is used, has the (lesser) length

$l\sqrt{1-\frac{v^2}{c^2}}$. This is the shortening of moving bodies introduced by Lorentz and Fitzgerald in their explanation of the Michelson experiment.

Here it follows as a consequence of our general postulates. According to the principle of relativity, this shortening cannot happen only with the movement of rigid bodies *relative* to K, but must rather occur with the movement relative to any admissible coordinate system. Thus, it is easily shown from (5), or the inversion of (5), that a rod oriented in the x-direction and relatively at rest in K, with length l in K, has, when seen from K', the length $l\sqrt{1-\frac{v^2}{c^2}}$. Therefore, one can say that either one of these two rods moving lengthwise relative to the other is—when seen

[p. 12] from the other one—shortened.

Next, we consider a unit clock permanently at rest in the origin of K', when its periods are characterized by the values

{2} $$t' = 0;\ t' = n\ (n = \text{integer})$$

From the first and fourth equation in (5) one finds

$$t = \frac{n}{\sqrt{1-\frac{v^2}{c^2}}}$$

Since $\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$ is larger than 1 whenever v is nonzero, it follows that a clock moving relative to a coordinate system—when viewed from this coordinate system—runs more slowly than if it were not moving.

These two consequences concerning the behavior of moving bodies and clocks are, in principle, testable by experiment. Because of considerable practical difficulties, none of these consequences has so far been tested experimentally, the reason being that in nearly all practically accessible cases $\frac{v^2}{c^2}$ is a practically vanishing quantity compared to the unit.

Considering the fundamental importance of the Lorentz transformation for the theory of relativity, we have to add that the principle of the constancy of the speed of light and the principle of relativity alone are not yet sufficient to derive the trans-

formation equations (5). The conditions that have to be added are such that we would not give them up without need; these assumptions are

a. Homogeneity of space. The behavior of measuring rods and clocks does not depend upon the locations where they are in space nor upon the time, but solely upon the manner in which they are moving. From this one can conclude that x', y', z', t' are linear functions of x, y, z, t.
b. Isotropy of space. The behavior of measuring rods and clocks is independent of the choice of the direction of movement.
c. Independence of measuring rods and clocks from their past history.[22]

Finally, we note that it is unessential that the spatial orientation of the coordinate systems K and K' be chosen according to fig. 2. Abandoning this unessential assumption, one arrives at somewhat more general transformations which, however, still satisfy conditions (6) and (6a). All these transformations are called, in a wider sense, Lorentz transformations; like equations (5), they are linear in the coordinates.

Equations (5) have, so far, found their most direct confirmation—as Herr Laue was the first one to notice—in the outcome of the Fizeau experiment (see [2]).[23] [p. 13]
Namely, imagine that the pipe in the experiment is at rest relative to K, its axis is oriented in the x-direction, but the fluid is at rest relative to K'. Then light propagation relative to K' occurs according to the equation

$$x' = Vt'.$$

Introducing the coordinates x and t into this equation by means of (5), one first gets

$$x - \mathrm{v}t = V\left(t - \frac{\mathrm{v}}{c^2}x\right),$$

and with slight rearranging and neglecting terms of the order $\frac{\mathrm{v}^2}{c^2}$, one finds the equation

$$V' = \frac{x}{t} = V + \left(1 - \frac{1}{n^2}\right)\mathrm{v}, \qquad \{3\}$$

which we mentioned above as the result of experience. The result, therefore, follows from the basic principles of the theory of special relativity without any analysis of the optical process.

12. *The Heuristic Meaning of the Lorentz Transformation*[24]

The principle of special relativity demands the equivalence of all those coordinate systems that are in a rectilinear, uniform, rotation-free motion relative to *one* (admissible) coordinate system; the same laws of nature shall be valid in reference to all these coordinate systems. On the other hand, it follows from our two principles that, between the coordinates of any two of these admissible coordinate systems, there obtain those relations we have called Lorentz transformations. Combining these two statements leads to this result:

The laws of nature must be such that the introduction of new coordinates by means of a Lorentz transformation does not change their form (covariance of the laws of nature under Lorentz transformations).

Maxwell-Lorentz electrodynamics satisfies this condition; Newtonian mechanics, however, does not. Therefore, the latter had to be modified in order to yield to this requirement. This could be accomplished without great difficulties; for example, it turned out that Newton's equations of motion for a pointlike mass,

$$\frac{d}{dt}(m\mathfrak{q}) = \mathfrak{K},$$

which say that the change rate of momentum equals the force, has to be replaced by

$$\frac{d}{dt}\left(\frac{m\mathfrak{q}}{\sqrt{1-\frac{q^2}{c^2}}}\right) = \mathfrak{K}.$$

[p. 14] In these equations, $\mathfrak{q}$ denotes the vector, q the amount of the velocity, and $\mathfrak{K}$ the force acting upon the material point. The two equations are not noticeably different so long as the velocity q is small compared to the speed of light. With Newton, $m\mathfrak{q}$ expresses the momentum (impetus) of the mass; according to the theory of relativ-

ity, it is $\frac{mq}{\sqrt{1-\frac{q^2}{c^2}}}$. The latter quantity approaches infinity when q approaches the speed of light; the first quantity does not. This is connected to the fact that, according to the theory of relativity, it is impossible, however large the forces or for however long they act, to raise the velocity of a material point ⟨body⟩ up to the speed of light (or beyond it). The speed of light, quite generally, plays within the theory the role of a physically infinitely large and insurmountable velocity; i.e., for projectiles, signals (waves) etc. This is already seen by glancing at the Lorentz transformation and the consequences previously drawn from it about measuring rods and clocks.

For the energy $\langle L\rangle E$ of a moving mass point, one finds the expression

$$\langle L\rangle E = \frac{mc^2}{\sqrt{1-\frac{q^2}{c^2}}} \tag{7}$$

or, developed in powers of q^2,

$$\langle L\rangle E = mc^2 + \frac{1}{2}mq^2 + \frac{3}{8}\frac{m}{c^2}q^4 \ldots. \tag{7a}$$

The second term in this expansion is the "kinetic energy" of classical mechanics. The third, fourth, etc. terms vanish against the second one when $\frac{q^2}{c^2}$ becomes negligible compared to the unit. The first term in (7a) deserves special attention. It has to be remembered that the energy of the masspoint itself does not follow from the equation of motion, so that the energy is defined only up to a constant which we omitted in (7). However, the first term mc^2 in (7a), to which the expression of E reduces in case $q = 0$, is formally so closely connected to the terms following (as a glance upon [7] shows) that we are inclined to attribute physical meaning to it. We can look at mc^2 as the energy of the masspoint when it is at rest ($q = 0$).

This interpretation receives mighty support from a theoretical investigation [p. 15]
based upon the following consideration.[25] According to the theory of special relativity, the theorem of the conservation of energy must hold not only relative to *one* coordinate system K but also relative to a system K' in uniform motion relative to K. From this one can derive—without here detailing how—the theorem:

The inertial mass m of a body increases by $\frac{E}{c^2}$ if one adds (in a state of rest) the energy E (e.g., in the form of heat, chemical energy, etc.).

This result, which is derivable with complete rigor from the premises of the theory (in connection with Maxwell's electrodynamics) states that *at least part* of the inertial mass of a body consists of energy. There cannot be much doubt that the mass of a body is nothing else than latent energy.

This reflection also shows that the theorem of the conservation of mass can no longer claim a place independent of the theorem of the conservation of energy.

13. *Theory of Special Relativity and Ether*

It is obvious that an *ether at rest* has no place in the theory of relativity. Because, if two systems K and K' are of completely equal value for the formulation of the laws of nature, then it is inconsistent to put a concept into the foundations of a theory such that it distinguishes one system over all the others. After all, if one assumes an ether that is at rest relative to K, it is moving relative to K', a feature that does not agree with the equivalence of the two systems.

For this reason, my opinion in 1905 was that one should no longer talk about the ether in physics. But this judgment was too radical, as we shall see in the following considerations on the theory of general relativity. Rather, it is still permissible to assume a space-filling medium whose states may be imagined as electromagnetic fields (and perhaps also as matter). But it is not permissible to attribute to this medium states of motion in every point, like in an analogy to ponderable matter. This ether must not be imagined as consisting of particles whose identity could be traced in time.[26]

14. *Minkowski's Method*[27]

[p. 16] In an era of classical dynamics it would have been idle play to combine time and space into a four-dimensional continuum. Because for that era, the chronological order of events was independent of everything spatial, i.e., independent of the choice of the coordinate system. In the theory of relativity, however, time is deprived of its separate existence; the Lorentz transformation demonstrates that spatial and chronological coordinates depend on each other according to the state of

motion of the coordinate system. ⟨The separate role of the time coordinate vis-à-vis the space coordinate vanishes.⟩ The mutual dependence of the spatial and the chronological necessarily leads to an amalgamation of space and time into a four-dimensional ⟨total⟩ continuum.

In addition, Minkowski also found that this four-dimensional continuum, which he called "world," has a deep-rooted formal kinship with the three-dimensional continuum of Euclidean geometry. We furthermore recognized that this kinship can be used as a basis for a method of direct formulation of equation systems (mathematically formulized laws of nature) that satisfy the principle of relativity, i.e., are covariant under Lorentz transformations. For the non-mathematician, it is not easy to grasp these connections. But still I will try to sketch their essence.

The structures of Euclidean geometry as well as those of theoretical physics have an existence independent of the spatial orientation of the Cartesian coordinate system. ⟨This corresponds to⟩ If I have formulated any relation using orthogonal coordinates x_1, x_2, x_3 of the coordinate system K, then this relation must present itself in an exactly corresponding form if I use a differently oriented orthogonal coordinate system K' (with the coordinates $x_1', x_2' x_3'$). Example: If R_1, R_2 are the radii of two spherical surfaces with x_1, x_2, x_3 and x_1^*, x_2^*, x_3^*, respectively, the coordinates of their centers, then these spherical surfaces do not intersect in K if the following equation is satisfied:[28]

$$R_1 + R_2 < \sqrt{(x_1^* - x_1)^2 + (x_2^* - x_2)^2 + (x_3^* - x_3)^2}. \qquad (8)$$

With respect to the coordinate system $K'(x_1', x_2', x_3')$, the condition of non-intersection must be the exactly corresponding one, that is,

$$R_1 + R_2 < \sqrt{(x_1^{*\prime} - x_1')^2 + (x_2^{*\prime} - x_2')^2 + (x_3^{*\prime} - x_3')^2}. \qquad (8a)$$

On the other hand, given the relative position of both coordinate systems, there are certain relations between the K-related coordinates x_1, x_2, x_3 and the K'-related coordinates x_1', x_2', x_3' of the same point P such that the x' can be expressed by means of the x (and vice versa the x by means of the x'). We want to call these [p. 17]
relations a "Euclidean transformation." If we express by means of a Euclidean transformation the coordinates x_1, x_2, x_3 and x_1^*, x_2^*, x_3^* in (8) through the $x_1', x_2' x_3'$ and $x_1^{*\prime} x_2^{*\prime} x_3^{*\prime}$, then we must obtain the condition of intersection of the two spherical surfaces in reference to K', i.e., we must get equation (8a). This

is what we want to express when we say: the condition of intersection is covariant under Euclidean transformations.

What has been said here in relation to conditions of intersection of two spherical surfaces must hold quite generally for every mathematically formulated law that is expressed on the basis of Euclidean geometry. All statements and laws must be covariant under Euclidean transformations in order to be meaningful from the point of view of Euclidean geometry. ⟨This is a general condition that all geometric and physical relations must satisfy.⟩

However, the Euclidean transformations, which, as we have said, have authoritative meaning for the construction of geometric and physical relations, are characterized in the most simple manner by the following condition. Euclidean transformations are those that satisfy (identically) the equation

$$dx_1^2 + dx_2^2 + dx_3^2 = dx_1'^2 + dx_2'^2 + dx_3'^2 . \qquad (9)$$

This relation says geometrically that the square of the distance (measured with the unit measuring rod) between two infinitesimally close points in space expresses itself by coordinate differences and in an identical manner for both coordinate systems, specifically by means of the Pythagorean theorem.[3][29] This distance is a quantity independent of the coordinate system, an "invariant," its square is the "fundamental invariant of Euclidean transformations."

Geometry and physics teach, especially in vector theory, how systems of equations are to be formed such that they are covariant under Euclidean transformations.

Again we return to the theory of special relativity. According to this theory—as we have seen—the laws of nature are structured so that they are covariant under Lorentz transformations (mathematical expression of the theory of special relativity). However, these transformations are characterized generally by satisfying identically the equation

$$dx^2 + dy^2 + dz^2 - c^2dt^2 = dx'^2 + dy'^2 + dz'^2 - c^2dt'^2 . \qquad (6a)$$

It strikes the eye how similar this condition is to the one for the Euclidean trans-
[p. 18] formations (9). The similarity becomes more complete when we introduce in place of the time coordinate t the imaginary coordinate $\sqrt{-1}ct$. Replacing x, y, z, t by x_1, x_2, x_3, x_4, putting

[3] This shows the fundamental importance of this theorem of our geometry.

$$x = x_1 \ , \ y = x_2 \ , \ z = x_3$$

$$\sqrt{-1}ct = x_4 ,$$

changes (6a) into[30]

$$dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 = dx_1'^2 + dx_2'^2 + dx_3'^2 + dx_4'^2. \qquad (10)$$

In a formal sense, this is the most simple condition to characterize the generalized Lorentz transformation.

Comparing (10) to (9), one realizes the complete mathematical analogy existing between the Lorentz transformation and the Euclidean transformation. The Lorentz transformation is a Euclidean transformation in a four-dimensional space-time continuum, provided one makes use of an imaginary time coordinate.[31]

Long ago, physics and geometry derived in a graphically vivid manner, by means of the theory of vectors, the formal qualities which the laws of nature must show in order to be covariant under purely spatial transformations (9). An analogous extension of these formal constructions into the four-dimensional provides us with these structures and equations that are covariant under generalized Lorentz transformations (10). By recognizing this, Minkowski considerably simplified the application of the theory of special relativity. By his method one is able to judge directly whether or not systems of equations do comply with the requirements of relativity, without having to carry out a transformation of them.

However, it has to be emphasized that the equivalence of the time coordinate x_4 with the spatial coordinates x_1, x_2, x_3 is only a formal and not a physical one, which is obvious from what has been said above.

As is obvious from (9), the entire conceptual system of Euclidean geometry can be built upon the fact that the square of the spatial distances

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 \qquad (11)$$

is a quantity independent of the choice of coordinates where ds is a quantity measurable with a ⟨unit⟩ measuring rod. In a corresponding way, it is of fundamental importance for the theory of special relativity that the quantity [p. 19]

$$d\sigma^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2 \quad (12)$$

is measurable by means of measuring rods and clocks and is independent of the choice of coordinates (inasfar as this choice is free). $d\sigma$ is the fundamental invariant of the system and is called the elementary distance of neighboring points of the space-time continuum because of its formal analogy with the fundamental invariant ds of Euclidean geometry.

But, since x_4 is imaginary and therefore dx_4^2 negative, there is, aside from the number of dimensions, a deep-rooted formal difference between the invariant ds of Euclidean geometry and the invariant $d\sigma$ of the theory of special relativity. Namely, ds^2 is always positive as long as the adjacent space points P and P'—to which ds is associated—do not coincide. However, when P and P' are neighboring space-time point-events ("world points" in Minkowski's terminology), then one must distinguish three possibilities as to their relative position, all of them independent of the choice of the coordinate system:

$$d\sigma^2 < 0$$
$$d\sigma^2 > 0$$
$$d\sigma^2 = 0.$$

The coordinate system for the first case can be chosen such that only dx_4 differs from zero while dx_1, dx_2, dx_3 vanish; the distance of events $\overline{PP'}$ is then called "timelike" and the distance $d\sigma$[32] of $\overline{PP'}$ can be directly measured by a clock that is at rest relative to the coordinate system chosen. For the second case, the coordinate system can be chosen such that the pointlike events P and P' occur simultaneously relative to it. $\overline{PP'}$ is then called "spacelike"; the associated $\sqrt{-d\sigma^2}$ can be measured by a measuring rod that is at rest relative to the coordinate system chosen.

The third case, which is a limiting case, is physically characterized by the fact that the two pointlike events (world points) can be connected by a vacuum light signal.

Therefore, the analogy existing between the properties of the four-dimensional "world" of the theory of special relativity and the "space" of Euclidean geometry is only a mathematically formal one, not a physical one.

II. The Theory of General Relativity

[p. 20]

15. *The Basic Idea of the Theory of General Relativity in Its Original Form*

The following idea on Faraday's magneto-electric induction—so far not mentioned—played a leading role for me when I established the theory of special relativity.[33]

According to Faraday, during the movement of a magnet relative to an electric circuit, an electric current is induced in the latter. It is irrelevant whether the magnet or the conductor is moved; what counts only is the relative motion. But according to the Maxwell-Lorentz theory, the theoretical interpretation of the phenomenon is very different for the two cases:

If the magnet is moved, there exists in space a magnetic field variable with time, which, according to Maxwell, forms closed lines of an electric force, i.e., a physically real electric field; this electric field then puts the movable electric masses in the conductor into motion.

However, no electric field arises if the magnet is at rest and the circuit is moved; instead, the current in the conductor is created because the electricities moving with it due to the (mechanically forced) movement relative to the magnetic field suffer an electromotive force, which Lorentz introduced hypothetically.

The idea that these two cases should essentially be different was unbearable to me. According to my conviction, the difference between the two could only lie in the choice of the point of view, but not in a real difference ⟨in the reality of nature⟩. As seen from the magnet, there was certainly *no* electric field; whereas seen from the circuit there certainly was an electric field. Therefore, the existence of the electric field was a relative one, depending on the state of motion of the coordinate system used; and only the electric and magnetic fields *combined*, aside from the state of motion of the observer or coordinate system, could be granted a kind of objective reality. This phenomenon of magneto-electric induction forced me to postulate the principle of (special) relativity.[4][34]

When I was busy (in 1907) writing a summary of my work on the theory of special relativity for the *Jahrbuch für Radioaktivität und Elektronik* [Yearbook for Radioactivity and Electronics],[35] I also had to try to modify the Newtonian theory of gravitation such as to fit its laws into the theory. While attempts in this direction

[4] The difficulty to be overcome was originally in the constancy of the speed of light in a vacuum which, initially, I thought I would have to abandon. Only after years of groping did I realize that the difficulty rested in the arbitrariness of the basic concepts of kinematics.

showed the practicability of this enterprise, they did not satisfy me because they [p. 21] would have had to be based upon unfounded physical hypotheses. At that moment I got the happiest thought of my life in the following form:

In an example worth considering, the gravitational field has a relative existence only in a manner similar to the electric field generated by magneto-electric induction. *Because for an observer in free-fall from the roof of a house there is during the fall*—at least in his immediate vicinity—*no gravitational field.*[36] Namely, if the observer lets go of any bodies, they remain relative to him, in a state of rest or uniform motion, independent of their special chemical or physical nature.[5][37] The observer, therefore, is justified in interpreting his state as being "at rest."

The extremely strange and confirmed experience that all bodies in the same gravitational field fall with the same acceleration immediately attains, through this idea, a deep physical meaning. Because if there were just one single thing to fall in a gravitational field in a manner different from all others, the observer could recognize from it that he is in a gravitational field and that he is falling. But if such a thing does not exist—as experience has shown with high precision—then there is no objective reason for the observer to consider himself as falling in a gravitational field. To the contrary, he has every right to consider himself in a state of rest and his vicinity as free of fields as far as gravitation is concerned.

The experimental fact that the acceleration in free-fall is independent of the material, therefore, is a powerful argument in favor of expanding the postulate of relativity to coordinate systems moving nonuniformly relative to each other.

On the other hand, one can also start with a space that has no gravitational field. A material point in this space, when sufficiently distant from other masses, behaves free of acceleration relative to an inertial system K. However, if one introduces a *uniformly accelerated* coordinate system K' relative to K (uniformly accelerated parallel translation), then K' is no inertial system in the sense of classical mechanics or the theory of special relativity. Every mass point sufficiently distant from others is uniformly accelerated relative to K'. When seen from K, the acceleration of the system K' is of course the cause of the relative acceleration of the mass point relative to K'; and on the basis of classical mechanics, as understood up to the present day, it is the only possible interpretation. However, we can also view K' as an admissible system ("at rest") and attribute the acceleration of masses relative to [p. 22] K' to a static gravitational field that fills the entire space that is under consideration. This interpretation again is possible based upon the experimental fact that in a gravitational field (such as that relative to K') all bodies fall in the same manner.

[5] Of course, this consideration ignores the air resistance.

If we know the laws of nature with respect to a (gravitation-free) system K, then we can by mere transformation learn the laws relative to K', i.e., we learn about the physical properties of a gravitational field by means of a purely speculative method. At its basis here is the hypothesis that the principle of relativity also holds in reference to coordinate systems that are mutually accelerated to each other, and that the physical properties of space that rule relative to K' are completely equivalent to a gravitational field (hypothesis of equivalence).[38]

The generalization of the principle of relativity, therefore, points to a speculative way of investigating the properties of the gravitational field.

Because all bodies in a gravitational field have the same fall, a stimulus arose that pointed with irresistible force toward a generalization of the principle of relativity. ⟨Consequently, it is necessary to point out that this result (of the equivalence hypothesis) is supported with extraordinary precision, in particular by the tests made by Eötvös. This is based upon the following consideration.⟩ This experimental fact can also be phrased in a second especially remarkable form. According to Newton's law of motion, the fall of a body occurs according to the equation

$$\begin{aligned}&(\text{inertial mass}) \times (\text{acceleration in fall})\\&= (\text{gravitational force of the earth}).\end{aligned}$$

On the other hand,

$$\begin{aligned}&(\text{gravitational force of the earth})\\&= (\text{intensity of the gravitational field}) \times (\text{gravitational mass}).\end{aligned}$$

In these equations "inertial mass" means the mass that is responsible for the inertial reaction of the body, "gravitational mass" is the constant responsible for the influence of the gravitational field on the same body—two constants which by definition are completely independent of each other. From both equations together follows

$$\begin{aligned}&(\text{inertial mass}) \times (\text{acceleration in fall})\\&= (\text{gravitational mass}) \times (\text{intensity of the gravitational field}).\end{aligned}$$

In order to keep the experimentally confirmed law

$$(\text{acceleration in fall}) = (\text{intensity of the gravitational field})$$

valid, it must also be true that

$$\text{inertial mass} = \text{gravitational mass}.$$

[p. 23] The experimental fact of the same fall of all bodies therefore can, in the spirit of Newtonian mechanics, also be viewed as the equality of the inertial and gravitational mass, which from the point of view of Newtonian mechanics is by no means self-evident.

This theorem has been confirmed with extraordinary precision by the tests of Eötvös,[39] which are based upon the following. A body on the surface of the earth is under the influence of the gravitational force of the earth and of the centrifugal force of the earth's rotation. The first force is proportional to the gravitational mass, the latter to the inertial mass. The resultant of both forces is independent of the material only if the ratio of inertial and gravitational mass is independent of the material. Eötvös attached masses of different material to the ends of the horizontal balance beam of a torsion scale. In case of an incomplete proportionality of inertial and gravitational mass, the resulting forces acting upon the two masses could not be exactly parallel; i.e., there should have been a torsion moment acting upon the system when the balance beam was oriented in the east-west direction. The negative outcome was registered with such precision that the relative difference between inertial and gravitational mass had to be smaller than 10^{-7}.

16. *General Reasons for a General Postulate of Relativity* [40]

Classical mechanics and the theory of special relativity know of admissible coordinate systems (inertial systems) and nonadmissable coordinate systems. Relative to the first ones, the laws of nature (e.g., the law of inertia and the theorem of the constancy of the speed of light) are supposed to hold; relative to the latter, they do not. In vain one asks for an objective reason for the different quality of systems; and one is forced to explain them as an independent, very strange property of the space-time continuum. Newton was only very reluctantly content with this opinion (of "absolute space"), but he believed that in centrifugal effects he had an objective proof for it in hand.

However, E. Mach was the first to recognize the weakness of this argument. Perhaps it was not a physical quality of *space* that determined the inertial behavior of bodies; it could also be possible that inertia was not a reaction against a (conceptually empty) acceleration with respect to *space,* but rather against an acceleration *with respect to the rest of gravitating matter in the world.* Such a hypothesis appeared more satisfactory to Mach than the old concept of inertia, because it did not

attribute to space any ⟨independent⟩ mechanically distinct properties but rather, in principle, accepted all coordinate systems as equal. According to this interpretation, inertia also was an interaction between bodies just as is the case in Newtonian [p. 24] gravitation. It is true that this idea did not yet point to a rigorous (quantitative) treatment of the problem, and in essence the natural equality between inertia and gravitation—as laid out earlier (hypothesis of equivalence)[41]—remained hidden to Mach. But he was (after Newton) the first to vividly feel and clearly illuminate the epistemological weakness of classical mechanics.

It should by no means be claimed that the basically unsubstantiated preference of inertial systems over other coordinate systems constitutes an *error* of classical mechanics. The preference of certain sates of motion (namely, of inertial systems) in nature could be a final fact that we have to accept without being able to explain it (or reduce it to some cause). However, a theory in which all states of motion of coordinate systems are—in principle—equal has to be appreciated from an epistemological point of view as being far more satisfying. For the following consideration we want to use this equivalence as a basis under the name of "general ⟨postulate⟩ principle of relativity."[42]

17. *Some Consequences of the Equivalence Hypothesis*

We consider a space-time domain in which, under suitable choice of the coordinate system K, there is no gravitational field (relative to K); thus, K is an inertial system in the sense of classical mechanics. The laws valid in reference to K, e.g., the law of the propagation of light, can then be viewed as known. Now we introduce besides K a second coordinate system K' that is accelerated relative to K. There then is a gravitational field relative to K' due to the equivalence hypothesis. Since it is possible to establish the course of natural processes in reference to K' by a mere transformation from K to K', one learns from this procedure what type of effect the gravitational field relative to K' has upon the natural processes under consideration.

A vacuum light ray proceeds rectilinearly and uniformly relative to K with velocity c. A simple, geometric consideration shows that this same light ray has a curvature relative to K' as soon as the light ray forms an angle with the direction of acceleration of the system. The gravitational force bends the light ray as if light were a catapulted heavy body.

This consequence is of great significance in a twofold way. First, it provides a [p. 25] criterion for the theory that is accessible to observation. Because , a simple calcu-

lation shows that a light ray passing closely by a celestial body of the size of the sun must suffer a deflection in the order of magnitude of one second of arc. It is, after all, this result the two English expeditions confirmed last year. However, it should be pointed out here that this simple consideration provides *numerically* only half of the actual value of deflection.[43] This is connected to the fact that the gravitational field in the theory of general relativity is not represented just by a vector field but rather by a formally more complicated structure, such that the transition from the parallel field—that obtains in reference K'—to the spherically symmetric field of a celestial body is not as simple as it might appear at first glance. This will be clarified later.

Second, this consequence shows that the law of the constancy of the speed of light no longer holds, according to the general theory of relativity, in spaces that have gravitational fields. As a simple geometric consideration shows, the curvature of light rays occurs only in spaces where the speed of light is spatially variable. From this it follows that the entire conceptual system of the theory of special relativity can claim rigorous validity only for those space-time domains where gravitational fields (under appropriately chosen coordinate systems) are absent. The theory of special relativity, therefore, applies only to a limiting case that is nowhere precisely realized in the real world. Nevertheless, this limiting case ⟨also⟩ is of fundamental significance for the theory of general relativity; because the fact from which we started out, namely that no gravitational field exists in the immediate vicinity of a free-falling observer, this very fact shows that in the vicinity of every world point the results of the theory of special relativity are valid (in the infinitesimal) for a suitably chosen local coordinate system.

This connection can be illustrated with a geometrical comparison from the theory of surfaces that turned out to be of decisive significance for the ⟨finding and⟩ implementation of the theory.[44] Metric relations in the plane are described by two-
[p. 26] dimensional Euclidean geometry; i.e., the constructions of Euclidean geometry can be executed by means of a compass and ruler such that a compass (set to a certain way) plays the role of a fixed distance and the ruler that of a straight line. If one wants to do geometry on a curved surface, e.g., on a sphere or an ellipsoid, instead of in the plane, then there are certain laws of construction with the compass (and a ruler osculating to the surface), but these laws are no longer expressed by those of the Euclidean geometry of two dimensions. In its geometrical properties, a small piece of the surface better approximates those of the plane the smaller it is ⟨an infinitesimally small piece of the curved surface approximates in its properties, without limit, those of an infinitesimal piece of the plane⟩. The Euclidean ⟨plane⟩ geometry of two dimensions applies to constructions with a compass and ruler on curved surfaces in infinitesimal domains. For Gaussian geometry of curved sur-

faces, this law is as fundamental as the validity of the theory of special relativity in infinitesimal domains of space-time is for the theory of general relativity.

18. *Influence of Gravitational Fields upon the Rate of Clocks. Redshift of Spectral Lines*

According to the theory of general relativity the equivalence of coordinate systems is not limited to those of nonuniform parallel translation (rotation-free motion). If the theory should eliminate the epistemological deficiency of classical mechanics that was mentioned above, then it must be capable of treating every coordinate system, whatever its state of motion relative to others may be, as "at rest," i.e., the general laws of nature must be expressed by identical equations relative to all systems, whichever way they are moving.

Again, we start with a domain that has no gravitational field relative to a coordinate system K. Therefore, K is an "inertial system" in the sense of classical mechanics. Next we introduce a second coordinate system K' that uniformly rotates[45] relative to K; we symbolize this systems as a circular disk uniformly rotating relative to K (see fig. 3). We now imagine two identically constructed clocks where one (U_1) is located in the center of the disk, the other (U_2) on its periphery, such that it partakes in the rotational motion. Judging the rate of these clocks from K (i.e., from the point of view of the nonrotating system), it is clear from a result of the theory of special relativity, which is valid in reference to K, that U_2 runs more slowly than U_1, because U_2 has a velocity relative to K (against the paper), whereas U_1 has not. That U_2 runs slower than U_1 should also be noticed by an observer sitting on the disk (say next to U_1); therefore U_2 runs more slowly than U_1 when judged from K'. [p. 27]

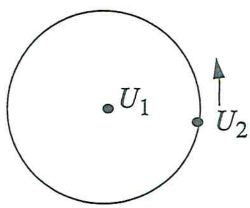


Fig. 3

According to the theory of general relativity, we can also view the system K' as "at rest." However, then we have to perceive the field of centrifugal forces existing relative to K' as a (real) gravitational field that acts upon all masses that are at rest relative to K' in proportion to their masses. (For the sake of completeness, we have to add that this gravitational field does not only consist of this ⟨gravitational⟩ centrifugal field, but it also has other components that act upon *moving* masses).[46] When judged from K' the two clocks are situated at different points in a gravitational field, and the latter is the cause that the two clocks run at different rates.

Quantitatively, one can conclude from this example that the rate of one and the same clock is proportional to $1 + \frac{\Phi}{c^2}$, where Φ is the gravitational potential at the location of the clock.

From this follows that a clock on the surface of a celestial body runs more slowly than the same clock when it floats (at rest) somewhere out in space or when it sits on the surface of a smaller celestial body. Every system is to be considered as a "clock" which by virtue of internal laws and periodically occurring processes is endowed with a specific frequency, that is, e.g., an atom that can emit or absorb a certain spectral line. When compared to the spectral lines generated by an element on earth the spectral lines generated or absorbed by the same element at the surface of the sun then must show a shift toward the red that is equivalent to a Doppler effect of about 0.6 km/sec. It still is doubtful if this necessary consequence of the theory is realized in nature; but according to recent investigations by two physicists at Bonn, Grebe and Bachem,[6][47] the reliable verification should be expected soon.

It should be noted that the law just mentioned can also be derived from the special case discussed above, that K' is in a uniformly accelerated translatory motion instead of being in rotation.

That identically constructed clocks, at rest relative to each other in a gravitational field, run at different rates at different locations is, in principle, of great principal
[p. 28] significance to us because from this fact it follows that in the theory of general relativity, time cannot simply be measured by identically constructed, suitably adjusted clocks that are at rest relative to each other, as is the case in the theory of special relativity. The direct physical meaning of time is thereby lost. In the following, we shall show that a corresponding situation exists with respect to spatial coordinates such that again we are confronted by the necessity to revise the physical interpretation of space and time.

19. *Invalidity of Euclidean Geometry in the Theory of General Relativity*

If an observer measures the circumference U and the diameter D of a circular disk (at rest) with a measuring rod practically infinitesimal compared to D, then the ratio $\frac{U}{D}$ from the two measurements is π $(3.14\ldots)$.This result can be taken as safe-

[6] Einstein left space here for a footnote, presumably to fill in the location of this reference later, but never did so.

ly established if the measured disk is at rest relative to an inertial system K (that is free of gravitational fields).

If an observer comoving with the previously considered circular disk makes the same test, then—we claim—the ration $\frac{U}{D}$ of the corresponding two results will be greater than π. This is shown by again viewing the entire process from the non-comoving coordinate system K. Judged from K, the measuring rod tangentially aligned at the periphery of the rotating disk is shortened (Lorentz contraction) due to its movement along this line; but the radially aligned measuring rod is not. According to the law of Lorentz contraction, one finds

$$\frac{U}{D} = \frac{\pi}{\sqrt{1 - \frac{v^2}{c^2}}},$$

where v is the rotation velocity at the edge of the disk.

From this it becomes obvious that on a rotating disk and, therefore, according to the hypothesis of equivalence also in a field of gravitation, *the laws of Euclidean geometry are not valid for the relative positioning of rigid rods*.[48] Specifically, it also becomes impossible to attribute physical meaning to a Cartesian coordinate system in the theory of general relativity because it will be impossible to construct a cubic lattice out of identical rods. We are therefore faced with the new difficulty that spatial and chronological coordinates cannot be physically defined by means of rigid rods and clocks as in the theory of special relativity. Thus, we face the difficult question: What is to take the place, in the theory of general relativity, of the Cartesian coordinate systems and time as defined by means of clocks and light signals in the theory of special relativity?

20. *Gaussian Coordinates. Riemannian Geometry* [p. 29]

As shall be shown in the following, the problem arising here has been solved in the geometric domain by Gauss and Riemann.[49] ⟨Their method can⟩

Considered as an axiomatic science, Euclidean geometry has at first nothing to do with objects of experience. Its theorems are consequences of so-called axioms, hence, in principle, are already implicitly contained in the latter. The axioms seem to be related to mere objects of thought that have nothing to do with objects of experience.

The abstraction of basic geometric concepts (straight lines, distances, etc.), from objects of experience whose abstracted images they are, serves a systematic interest but should not delude us; geometry, after all, is created in order to teach us about the behavior of things of daily experience. If there were no practically rigid bodies that can be brought into congruence with each other, we would not talk about congruence of lengths, triangles, etc. It is obvious that geometry obtains meaning for the physicist only by the fact that he associates natural things to these basic concepts, e.g., to the concept of distance between two marks affixed to a practically rigid body. In turn, Euclidean geometry becomes an experimental science in the proper sense due to this association, just like mechanics. Its theorem then can be confirmed or disproven in the same sense as those of mechanics. Our previously found result, that Euclidean geometry does not obtain on a rotating circular disk, is to be interpreted in this sense.

Euclidean geometry became important for physics primarily in the analytical form we owe to Descartes. This transformation of Euclid's system was possible through the discovery of the Cartesian coordinate system whose physical interpretation we have considered already. Its introduction (e.g., in plane geometry) is possible by covering the plane with a net of squares whose sides consist of identical rodlets that can be brought into mutual congruence. That means one can, beginning at an arbitrary point in the grid, completely characterize any other point in the grid by two numbers (coordinates). Thus, coordinates have an immediate physical meaning.

Gauss now posed the problem to establish in an analogous manner an analytical geometry on an arbitrarily given curved surface. This requires, first of all, characterizing all points on the surface by means of numbers (coordinates). But now it became obvious that there is no such quadratic grid to define Cartesian coordinates.
[p. 30] The reason being that on a curved surface the laws for the placement of rigid little rods are not provided by the Euclidean line segment geometry of the plane. Cartesian surface coordinates therefore do not exist on a curved surface (e.g., an ellipsoidal or spherical surface).

However, what still exists on a curved surface is the distance ds of neighboring points, as it can be measured by rigid measuring rods. Geometry on the curved surface must be based upon this concept, except that the simple relation $ds^2 = dx^2 + dy^2$, pertaining to Cartesian coordinate differentials and elementary distance, is dropped on curved surfaces.

The possibility of introducing Cartesian coordinates in the plane is based on the existence of a system of mutually orthogonal and parallel straight lines that can be numbered in sequence of distance. However, a curved surface has no ⟨completely adequate⟩ analog of parallel straight lines. Therefore, Gauss used for the definition

of coordinates, instead of two systems of parallel straight lines, two systems of arbitrary sets of curves with the sole requirement that only *one* curve of each set should go through each point of the surface. The curves of each set are numbered such that ⟨continuously⟩ neighboring curves get neighboring numbers. Then, two curves run through every point of the surface, one curve from each set, and their numbers are called its "coordinates" (x_1, x_2). This Gaussian coordinate system is—speaking graphically—nothing other than an arbitrarily deformed and stretched planar Cartesian coordinate system. It is by virtue of this bending that Gaussian coordinates no longer have any physical meaning whatsoever. They are not more than a completely arbitrary numbering of points of the surface so that, however, continuity is preserved.

Nevertheless, the distance ds between two neighboring points P_1, P_2 on the surface is expressable in a certain law of the coordinate differences dx_1, dx_2. After all, an infinitesimal piece of surface can always be viewed in first-order approximation as a piece of a plane; and there exists for the element a *locally Cartesian* coordinate system (X_1, X_2) such that

$$ds^2 = dX_1^2 + dX_2^2.$$

If one replaces these locally Cartesian coordinates with Gaussian coordinates (x_1, x_2), one obtains by a simple reflection

$$ds^2 = g_{11}dx_1^2 + 2g_{12}dx_1dx_2 + g_{22}dx_2^2. \qquad (13)$$

The g_{11} etc. are here functions of x_1 and x_2 which are determined, on the one hand, by the choice of the Gaussian coordinates, and on the other hand also by the choice of the surface whose geometrical laws we want to study. If the functions g_{11} [p. 31] etc. are known, then they also define the metric properties of the surface.

Equation (13) is the Gaussian generalization of the Euclidean-Pythagorean theorem. In case of a plane and for Cartesian coordinates $(g_{11} = g_{22} = 1; g_{12} = 0)$, it takes the characteristic form

$$ds^2 = dx_1^2 + dx_2^2$$

of the Euclidean plane.

For the entire consideration, it is immaterial that the surface considered is part of the three-dimensional space and—when looked at in that space—is curved. Essential is only that we have before us a two-dimensional continuum whose laws of measurement in the infinitesimal are Euclidean but deviate from this norm in the finite domain. The metric behavior of the infinitesimal and rigid little rods on a

curved surface are—when seen two-dimensionally—non-Euclidean even though the little rod, in three-dimensional space, is supposed to play the role of a length in the sense of Euclidean geometry in three dimensions.

It is to be well noted that Euclidean or non-Euclidean behavior is not a property of the surface itself but rather one of certain measuring rodlets in reference to the surface. The disk-type surface considered, e.g., sub 19), is non-Euclidean for co-rotating measurement rodlets, but Euclidean for non-corotating ones. Geometric statements always refer to the possibilities for the placement of rigid bodies.

Riemann extended these considerations to manifolds of three or more dimensions. This is possible without difficulty if one considers that the entire analysis given above is independent ⟨and makes no use⟩ of the assumption that the surface considered is part of a three-dimensional Euclidean space. So let there be a three-dimensional space with the property, in reference to rigid rodlets of a certain kind, that the latter can only be arranged according to the laws of Euclidean geometry in the infinitesimal, but not so in the finite domain. Then there are (in the finite) no coordinates by means of which the elementary distance of neighboring points could be represented with rodlets in accordance with the formula

$$ds^2 = dx_1^2 + dx_2^2 + dx_3^2 .$$

However, if one characterizes the individual points of the space—under preservation of continuity—arbitrarily by means of three "Gaussian coordinates," then the distance squared expresses itself in the form

$$\begin{aligned} ds^2 = {} & g_{11}dx_1^2 + g_{22}dx_2^2 + g_{33}dx_3^2 + 2g_{12}dx_1dx_2 \\ & + 2g_{13}dx_1dx_3 + 2g_{23}dx_2dx_3 . \end{aligned} \tag{14}$$

[p. 32] The quantities g_{11}, etc., which themselves depend upon the coordinates x_1, x_2, x_3, completely describe the possibilities of placing "rigid" measuring rodlets into the space considered. The metric properties of all structures definable by our rodlets within this space must be derivable from them by pure calculation.

The theoretical formulas ⟨and results⟩ obtainable in this manner are of great generality in their formulation insofar as they hold for arbitrary coordinates and not just for a certain choice of coordinates (Cartesian coordinates) like those of Euclidean geometry. If one introduces instead of the original Gaussian variables x_1, x_2, x_3 arbitrary functions x'_1, x'_2, x'_3 of these quantities as new coordinates, one finds exactly corresponding formulas. Theorems formulated with the use of

general Gaussian coordinates are—in mathematical terminology—covariant under arbitrary transformations of the coordinates.

21. *Physical and Mathematical Content of the Principle of General Relativity*

After these more formal intermediary considerations, we pick up the thread of our main objective. We have seen how the extension of the principle of relativity upon non-acceleration-free relative movements of coordinate systems helped us to understand the essential equality of an inertial and gravitational mass. On the other hand, however, it turned out to be impossible to introduce into finite domains space-time coordinates such that the spatial coordinates could be measured with identically constructed measuring rods, and the timelike coordinates could be directly measured with clocks. There are no physical objects whatsoever to represent the straight line, whereupon, consequently, it becomes impossible to distinguish in a physically meaningful manner between rectilinear-orthogonal (Cartesian) and curvilinear coordinate systems. Therefore, we have here for the four-dimensional space-time continuum of physics a case that is precisely analogous to the geometrical problem of Gauss and Riemann.

But the analogy goes even further. Just as the metric behavior of infinitesimal pieces of a surface (or Riemannian space) is amenable to Euclidean geometry, such that there exist in the infinitesimal locally Cartesian coordinate systems in which the distance ds, measured with a measuring rod, can be expressed in (locally) directly measurable coordinates (X_1, X_2) according to the Pythagorean formula

$$ds^2 = dX_1^2 + dX_2^2;$$

exactly in the same way there exist everywhere in the space-time world of the theory of general relativity local coordinate systems where the simple metric relations of the theory of special relativity obtain. Just as in the theory of special relativity, the space-time coordinates are directly connected to results of measurements that [p. 33]
can be obtained with measurement rods and clocks; and, also, the Minkowski invariant $d\sigma$, given by the Pythagorean formula[50]

$$d\sigma^2 = dX_1^2 + dX_2^2 + dX_3^2 + dX_4^2 \qquad (15)$$ {4}

does exist. This formula is uniquely determined for all locations once a unit rod of measurement has been chosen.

Hence, it is clear that justice is done to the metric conditions that rule the four-dimensional world when we introduce general Gaussian coordinates x_1, x_2, x_3, x_4 into the four-dimensional world by perceiving the world as a four-dimensional Riemannian space in which the metric invariant $d\sigma$—which belongs to two neighboring space-time points—is physically uniquely defined by the generalized Pythagorean invariant

{4}
$$d\sigma^2 = g_{11}dx_1^2 + 2g_{12}dx_1dx_2 + \ldots + g_{44}dx_4^2. \qquad (16)$$

The coefficients $g_{11}, \ldots, g_{44}$ depending on x_1, x_2, x_3, x_4 not only determine the metric behavior of the world, i.e., the behavior of measuring rods and clocks, but also the phenomena of inertia and gravitation, as can be concluded from the hypothesis of equivalence.

For let us next look at a finite domain where the theory of special relativity is valid with sufficient approximation. With a suitable choice of the coordinate system K (inertial system) and measurement of time, the formula

$$d\sigma^2 = dx_1^2 + dx_2^2 + dx_3^2 + dx_4^2$$

is valid in such a subdomain. If we now introduce a new coordinate system K' (x'_1, x'_2, x'_3, x'_4) that is accelerated relative to K, then this is mathematically equivalent to the introduction of new space-time variables that are connected to the original ones by a nonlinear transformation. One finds by direct calculation that the
{4} metric invariant $d\sigma$ in the new system K' is represented by a formula like (16), whereby not all coefficients $g_{\mu\nu}$ are constant. On the other hand, we know from the hypothesis of equivalence that there is a gravitational field relative to K'.

We thus arrive at the result: the physical behavior of the space-time continuum
{5} is governed by 10 quantities $g_{11}, \ldots, g_{44}$ that determine the metric qualities (behavior of measuring rods and clocks) as well as the phenomena of inertia and gravitation.

[p. 34] The question of the mathematical formulation of the principle of general relativity is thereby decided. While the special theory of relativity demanded the covariance of the equations that express laws of nature under certain linear transformations of coordinates (Lorentz transformations), the general theory of relativity demands covariance under *arbitrary* transformations. In this theory, coordinates merely have the role of arithmetic parameters, devoid of any direct

connections to physical reality (except for the number 4 that is required by the nature of the space-time continuum).

22. *General Theory of Relativity and Ether*

There is no difficulty to include the laws of nature already known from the special theory of relativity into the wider framework of the general theory of relativity. The mathematical methods were completely at hand in the "absolute differential calculus" based on the research of Gauss and Riemann and further refined by Ricci and Levi-Civita.[51] It deals with the simple process of generalizing the equations from the special case of constant $g_{\mu\nu}$ to the case of $g_{\mu\nu}$ that are variable in space-time. In all of these generalized laws, the gravitational potentials $g_{\mu\nu}$ play the role which—in short—expresses the physical properties of empty space.

Again, "empty" space seems to be endowed with physical properties, that is, not physically empty as it appeared in the special theory of relativity. Therefore, one can say the ether has been resurrected in the theory of general relativity, even though in a ⟨newer⟩ more sublime form. The ether of the general theory of relativity differs from the one in old optics by not being a substance in the sense of mechanics. Not even the concept of motion can be applied to it. Furthermore, it is by no means homogeneous, and its ⟨structure⟩ state has no independent existence but rather depends upon the field-generating matter. Since the metric facts can no longer be separated in the new theory from the physical facts "proper," the concepts of "space" and "ether" flow into each other.[52] Since the properties of space appear to be conditioned by matter, space is no longer a precondition for matter in the new theory. The theory of space (geometry) and time can no longer be treated before physics proper or developed independently of mechanics and gravitation.

23. *The Field Law of Gravitation* [p. 35]

The most important problem of the general theory of relativity concerns the law of gravitation. This law found no place in the special theory of relativity because the potentials of gravitation $g_{\mu\nu}$ are replaced there by certain constants. Nevertheless, the idea of relativity leads to the solution of the problem of gravitation. For short ⟨and simplicity⟩ we shall call a domain ⟨or space⟩ where the special theory of relativity is valid a "Galilean space."

The field equations of gravitation we are looking for must be satisfied in a Galilean space, also for every Gaussian coordinate system to which we might refer it.

Furthermore, we know from experience that the gravitational field is determined by its mass, i.e., according to the special theory of relativity by the energy of matter.

If one adds the condition that the desired field equations should contain—like the Newton-Poisson equations of the classical theory—no derivatives of potentials of higher than second order, and those should be linear,[7][53] then the field equations are uniquely prescribed by the theory such that their formulation becomes possible in a mathematically deductive manner.[54]

A. Einstein

Translator's Notes

{1} In the second quadruple, dt has been corrected to dt'.

{2} $x' = 0$ has been corrected to $t' = 0$.

{3} The term $+\frac{1}{n^2}$ has been corrected to $-\frac{1}{n^2}$.

{4} Correct equation numbers have been inserted.

{5} The phrase "by 10 quantities" implies that $g_{ik} = g_{ki}$.

[7] From a mathematical point of view, this is the simplest possibility imaginable.